

Chapter V

STUDY SELECTION AND PARTICIPATION BIAS

1. Introduction

The main emphasis in the design and conduct of any epidemiologic study is comparability of the groups under study (Monson, 1980), and the strength of epidemiologic inference is directly associated with group comparability. In this study, Ranch Hand and comparison group comparability was assured by design since strict criteria were used to define the exposed (Ranch Hand) and the nonexposed (comparison) cohorts and since replacement comparisons were to be matched to original comparisons by perception of health. The cohorts were matched on the variables of age, race, and occupation group to minimize confounding and assure comparability in these variables. Within the nonexposed cohort, however, 4 subgroups resulted from the original match, the removal of ineligibles, replacement for noncompliance, the termination of the questionnaire and physical examination contracts, and the lack of data to match replacements to original comparisons. These groups are: original comparisons (O), shifted comparisons (S), replacement comparisons (R), and those replacement comparisons questioned by experienced Air Force interviewers (A). Because of logistic limitations, scheduling opportunities differed somewhat for each of these groups. Since compliance with this study was voluntary, the occurrence of differing scheduling options could have resulted in inadvertent selection bias (Cook and Campbell, 1979). The purpose of this chapter is to present the factors known to influence study participation, describe and analyze the responses of the Ranch Hand and the comparison groups to the opportunity to participate and to assess the potential bias of differential compliance. The analytic, and inferential implications of self-selection and potential participation bias will also be discussed. Participation is described in terms of location and compliance. A total of 1208 Ranch Hands and 1669 comparisons were the potential participants in this morbidity study.

2. Factors Known to Influence Study Participation

The study protocol estimated that 65% of the Ranch Handers would participate in the questionnaire and that 60% of these subjects would also participate in the physical examination. One major reason for these low estimates was the recognition of the negative influence of employment in flying occupations on compliance to physical examination. This negative influence was reinforced in the press and the subsequent advice of the Airline Pilots Association to their members not to participate in this study. This difficulty was anticipated by the principal investigators and is discussed in section VIII of the study protocol. Table V-1 presents a list of factors that could affect study participation. Those components of each factor that are considered in the study protocol for data collection are identified with an asterisk.

Table V-1

FACTORS POTENTIALLY AFFECTING STUDY PARTICIPATION

<u>Factors</u>	<u>Components</u>
Health Bias	*Self perception Current Use Long Term Care Abortion Pattern *Absenteeism *Current Medications Fertility History Current Family Health Familial History Severity of Past Disease Pending Retirement Bias Death
Logistic Factors	*Time Away From Family *Time Away From Job Distance to Exam Site *Income *Active Pilot (FAA)
Other Factors	Flying Status (USAF) Officer/Enlisted Age Race Current Status: AD/Sep Stipend Employment Status Dissatisfaction with Military
"Operational Factors"	Manner of Study Contact Scheduling Window Interviewer Bias
Publicity Bias	Motivational Bias Compensation Bias

The factors and the outlined components of each factor suggest the complexity of the compliance/noncompliance decision made by each study participant, Ranch Hander or comparison. The importance assigned to each component by the individuals in the Ranch Hand and comparison groups is most likely not equivalent. The Ranch Hand group was actively encouraged by the Ranch Hand Association to participate while no such organization exists for the comparison group.

3. Location

Mailing addresses for each study subject were determined through multiple military and civilian sources. Study subject location was initially identified by a certified mailing to these addresses. Current mailing addresses could not be identified for the nonlocatable population. Two-tenths percent of the Ranch Hand and 0.5% of the total comparison group were nonlocatable. This is well above the 99% location rate estimated in the study protocol. Table V-2 presents the counts of the located/nonlocated population by Ranch Hand and type of comparison.

Table V-2

COUNTS AND PERCENT OF LOCATABLE/NON-LOCATABLE
ALIVE STUDY SUBJECTS BY RANCH HAND AND
NATURE OF THE COMPARISON GROUP

	<u>Ranch Hand</u>	<u>Comparison</u>			
		<u>Original</u>	<u>Shifted</u>	<u>Replacements*</u>	<u>Total</u>
Locate	1206 (99.8%)	1023 (99.7%)	212 (100%)	425 (98.6%)	1660 (99.5%)
NonLocate	2 (0.2%)	3 (0.3%)	-	6 (1.4%)	9 (0.5%)
	1208	1026	212	431	1669

*Includes those individuals interviewed by USAF interviewers (A).

The two unlocated Ranch Hand individuals were separated from the military, and both had been nonflying enlisted personnel when on active duty. One was Black and the other was non-Black. Three of the 9 unlocatable comparisons were in the originally selected cohort. These 3 individuals were separated from the military, enlisted when on active duty (1 was a flying enlisted while the other 2 were nonflying enlisted) and all were non-Black. The locate algorithm was not completed on the replaced comparison "cannot-locate" population. Five of these 6 individuals were non-Black. The Black individual was separated and had served in an enlisted nonflying capacity. One other separated nonflying enlisted individual was non-Black. The remaining 4 replaced nonlocated comparisons were non-Black pilots. Two of these were separated, 1 was on reserve status and the other was retired. Overall, nonlocation did not impact data collection in this study. The 11 nonlocatable subjects are assumed to be alive and location will be attempted for the follow-up phases of the study. The replacement comparison group nonlocatable rate of 1.4% is of borderline significance when contrasted with the rate in the originally selected group ($P = 0.06$). This test was performed on the proportions using the normal approximation to the binomial. This difference was a result of the termination of the questionnaire contract prior to completion of the examination process. The names of 3 of the 6 replacement individuals were not sent to the questionnaire contractor while the 3 others were sent only 1 month prior to contract termination. The

replacement strategy as designed in the study protocol could not be implemented due to termination of the questionnaire contract prior to the completion of the physical examination contract.

4. Study Participation: Compliance

Study participation was characterized as being either fully compliant (FC) (completed the physical examination and the questionnaire); partially compliant (PC) (completed only the questionnaire) or noncompliant (NC) (refused the physical examination and the in-home questionnaire). Within the noncompliant group are those who completed an abbreviated telephone questionnaire. Figure V-1 shows that of the 1206 locatable Ranch Handlers alive at the initiation of the morbidity study, 1045 were fully compliant to the physical examination and an additional 129 completed the questionnaire but refused the physical examination. Ten of the 32 noncompliant Ranch Handlers completed the telephone questionnaire.

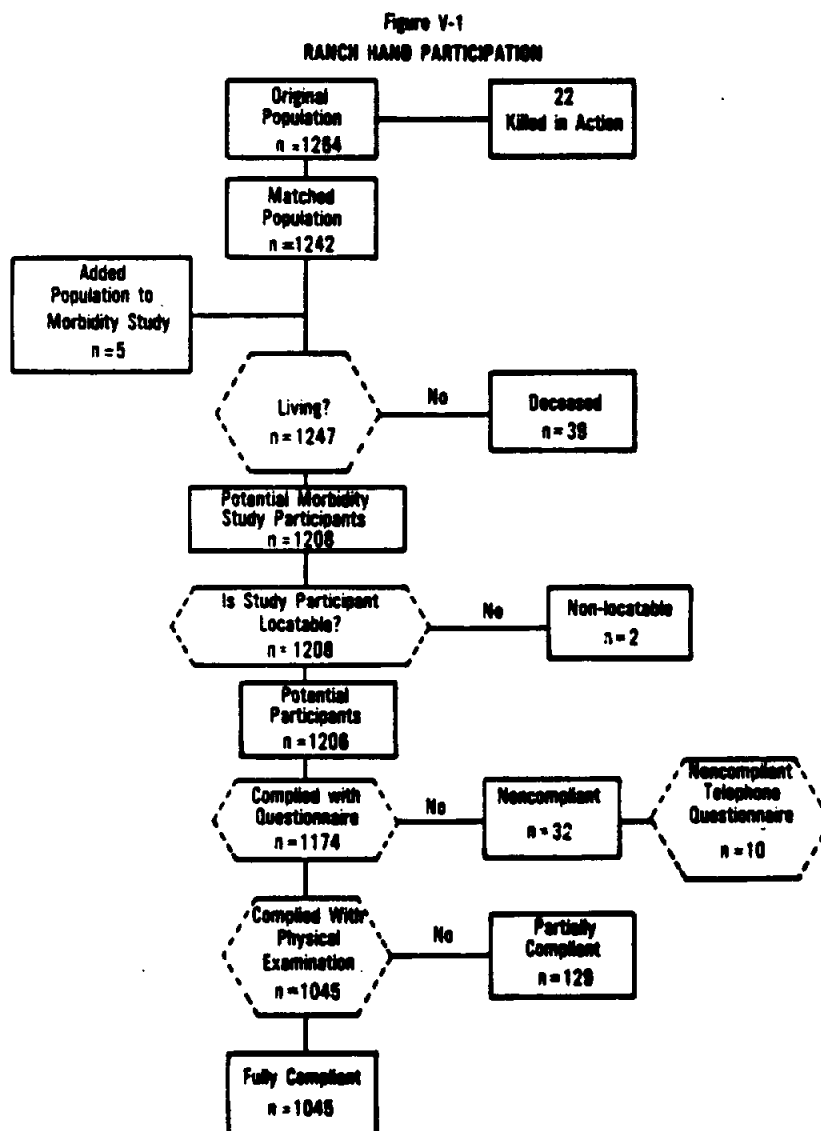
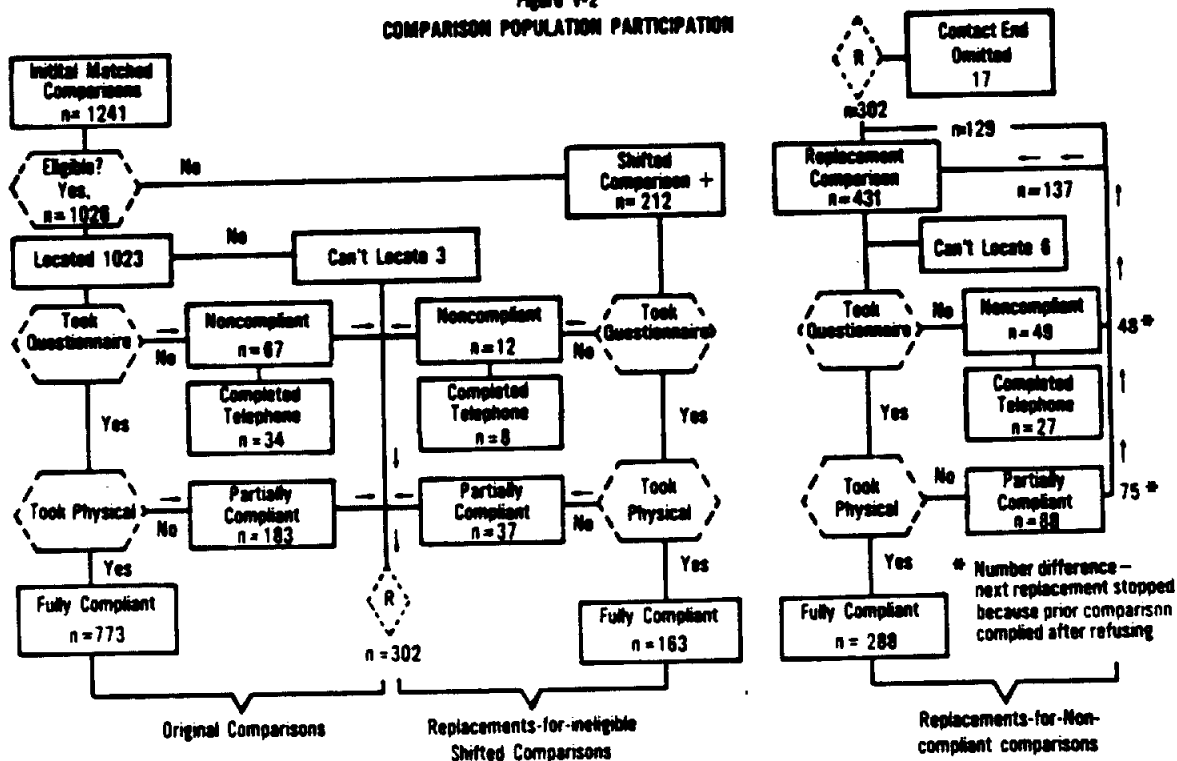


Figure V-2 describes the compliance patterns for the original, shifted and replaced comparison population. Of the 1023 locatable eligible original comparisons, 773 were fully compliant, 183 were partially compliant and 67 were noncompliant. Thirty-four of the noncompliant individuals completed the short telephone questionnaire. Thirty-four of the noncompliant individuals completed the short telephone questionnaire.

Figure V-2
COMPARISON POPULATION PARTICIPATION



Data collected by the noncompliant telephone instrument was delivered to the United States Air Force in written format following the implementation of the replacement strategy. The telephone questionnaire was not administered to the noncompliant replacement candidates prior to selection for the study, and therefore, the data necessary to match the original and replacement comparisons by similar perception of health status was not available (Lathrop, 1982). The next living individuals in the designated matched sets were selected as replacements. The data collected in the noncompliant instrument will be discussed in future publications.

Figures V-1 and V-2 are summarized in Table V-3, in which Ranch Hand and comparison participation is presented.

Table V-3

FULL, PARTIAL, NONCOMPLIANCE OF THE RANCH HAND AND COMPARISON
POPULATION BY NATURE OF THE COMPARISON GROUP, i.e.,
ORIGINAL (O), SHIFTED (S), REPLACED (R), AIR FORCE INTERVIEWERS (A)

	<u>RH</u>	<u>Comparisons</u>				<u>Total</u>
		<u>O</u>	<u>S</u>	<u>R</u>	<u>A</u>	
Fully Compliant (FC)	1045*	773	163	258	30	1224
Partially Compliant (PC)	129	183	37	88	-	308
NonCompliant (NC)	<u>32</u>	<u>67</u>	<u>12</u>	<u>49</u>	<u>-</u>	<u>128</u>
TOTALS	1206	1023	212**	395	30	1660

*4 individuals were interviewed at the Physical Examination site by USAF interviewers.

**3 Additional shifted comparisons were removed due to ineligibility identified following data collection.

The mean age of the population by compliance group is presented in Table V-4.

Table V-4

MEAN AGE OF THE RANCH HAND AND COMPARISON POPULATION BY NATURE
OF THE COMPARISON GROUP (O, S, R) AND TYPE OF COMPLIANCE (NC, PC, FC)

<u>Type Compliance</u>	<u>Ranch Hand Mean Age</u>	<u>Comparison Mean Age</u>		
		<u>O</u>	<u>S</u>	<u>R*</u>
Non-Black				
NC	41	41	39	40
PC	43	42	39	41
FC	44	45	43	41
Black				
NC	39	39	35	34
PC	39	43	39	38
FC	41	42	42	40

*Includes those individuals interviewed by USAF interviewers (A).

Table V-4 indicates that the noncompliant group is on the average younger than either the partially or fully compliant in both Black and non-Black strata. The compliant population is further described by race in Table V-5. This data is abstracted from Appendix XII, Occupational Category and Race of the Fully Compliant Population in Percent and Counts.

Table V-5

PERCENT FULLY COMPLIANT OFFICER/ENLISTED CATEGORIES BY RACE
RANCH HAND AND COMPARISONS (O, S, R)

		<u>Comparison</u>		
	<u>Ranch Hand</u>	<u>Original</u>	<u>Shifted</u>	<u>Replacements</u>
Non-Black				
Officers	85%	73%	78%	61%
Enlisted	88%	77%	77%	74%
Black				
Officers	67%	88%	*	*
Enlisted	90%	75%	69%	62%

* No individuals in this category.

This table suggests that Ranch Hand enlisted personnel complied at higher rates than officers and that Ranch Hand non-Black officers complied more than Black officers. The number of Black participants is very small and is therefore not included in the following analyses but is included in Appendix XII.

Appendix XVII was used to construct the data in Table V-6. Flying status is presented as flying/nonflying which includes both military and civilian information. Military status is categorized as active duty, retired, and separated/reserve.

Table V-6

PERCENT FULLY COMPLIANT OFFICERS BY FLYING STATUS AND MILITARY CATEGORY
(NON-BLACK ONLY)

	Ranch Hand n=372		Comparison					
			Original n=283		Shifted n=46		Replacements n=113	
	Flying	Non-Flying	Flying	Non-Flying	Flying	Non-Flying	Flying	Non-Flying
Active Duty (A)	77.8	96.3	58.9	76.2	87.5	75.0	57.9	88.9
Retired (R)	86.0	93.5	86.0	86.5	100.0	96.0	83.3	77.1
Separated/ Reserve (SV)	51.9	87.0	39.3	62.9	37.5	61.5	32.4	63.0

The flying separated/reserve category in this data set complied less than any other strata ($P < 0.01$), and flying status contributed significantly to the compliance decision ($P < 0.01$).

As illustrated in Table V-6, a complex set of interactions was involved in compliance. A log-linear model which was fitted to the three-way frequency table based on flying/military status, compliance, and group membership, revealed a three-way interaction ($P = .07$) in these data, rendering interpretations based on simpler models misleading. Since age and race are also related to flying/military status, tests of association between these factors and compliance need to be studied in the context of the many interactions present. These more complex relationships will be explored in future reports.

A summary of compliance is presented in Table V-7.

Table V-7

PERCENT OF THE STUDY POPULATION COMPLYING TO THE
QUESTIONNAIRE AND PHYSICAL EXAMINATION

	Ranch Hand	Comparison		
		<u>Original</u>	<u>Shifted</u>	<u>Replacements</u>
Questionnaire	97% (1174/1206)	92% (956/1023)	94% (200/212)	88% (376/425)
Physical Examination	87% (1045/1206)	76% (773/1023)	77% (163/212)	68% (288/425)

Ranch Hand personnel participated in the questionnaire at a rate higher than all comparison groups. This participation was 32% greater than the original protocol estimate of Ranch Hand compliance. Differential compliance to questionnaire did occur in the comparison groups with the original and shifted group complying 5% more than the replaced comparison group (unadjusted; $P=0.003$). Table V-7 shows that differential compliance also occurred between the Ranch Hand and the original comparison group in their compliance to physical examination (unadjusted; $P<0.001$) as well as within the comparison groups with the original and shifted comparison groups complying 8-9% more than the replaced group (unadjusted; $P<0.001$).

5. Noncompliance

The reasons given by study participants for noncompliance were compared. Appendixes XIII and XIV display all reasons given. These data were collected in a nonstandard manner by Louis Harris and Associates, the Kelsey-Seybold Clinic, and USAF personnel. The responses were then allocated to the categories presented in the appendix. They describe that the majority of the reasons given for noncompliance were "no time-no interest" and passive refusal. Table V-8 shows the percent of refusals in the Ranch Hand and comparison groups implying these disinterest reasons.

Table V-8

PERCENT OF REFUSALS CATEGORIZED AS REFUSALS FOR REASONS OF DISINTEREST

	<u>Ranch Hand</u>	<u>Comparison</u>		
		<u>Original</u>	<u>Shifted</u>	<u>Replacements</u>
Questionnaire	86%	67%	91%	49%
Physical Examination	50%	58%	54%	58%

These data indicate that the noncompliant replacement comparisons were passive refusals less often than were the other comparison groups. The percent refusals due to job commitment and confidentiality are described in Table V-9.

Table V-9

PERCENT OF QUESTIONNAIRE REFUSALS CATEGORIZED AS
JOB COMMITMENT AND CONFIDENTIALITY

	<u>Ranch Hand</u>	<u>Comparison</u>		
		<u>Original</u>	<u>Shifted</u>	<u>Replacements</u>
Job Commitment	-	3%	-	24%
Confidentiality/ Active Duty	<u>5%</u>	<u>14%</u>	-	<u>24%</u>
TOTAL	5%	17%	-	48%

Forty-eight percent of the replaced population stated that they refused to participate in the questionnaire because of a job commitment or the issue of confidentiality.

6. Scheduling Opportunity

The names of the Ranch Hand and original comparison groups were provided to the questionnaire contractor in November 1981. The contractor was given the shifted comparison population in April 1982 and the replacement population continued to be identified to the contractor through 15 Nov 1982. Physical examination scheduling was contingent upon completion of the questionnaire. Therefore, while the Ranch Handers and the original comparisons had 1 year to schedule and complete the study, the shifted comparisons had a maximum of 9 months, and the replacement comparisons were afforded a more limited scheduling opportunity.

Figure V-3

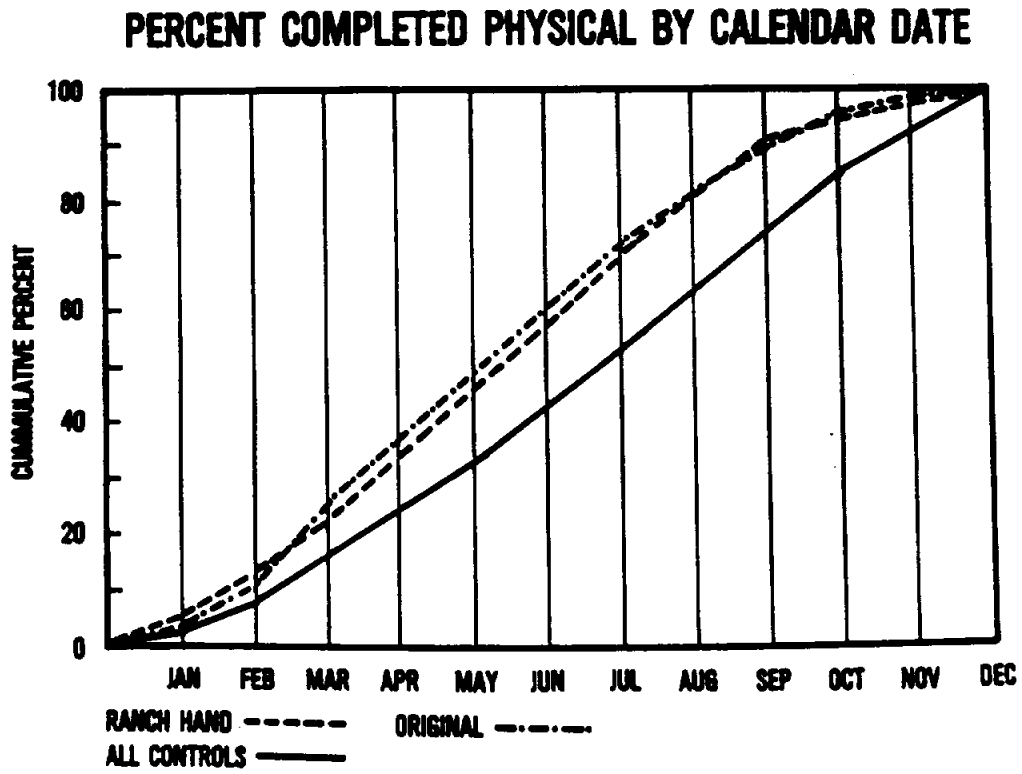
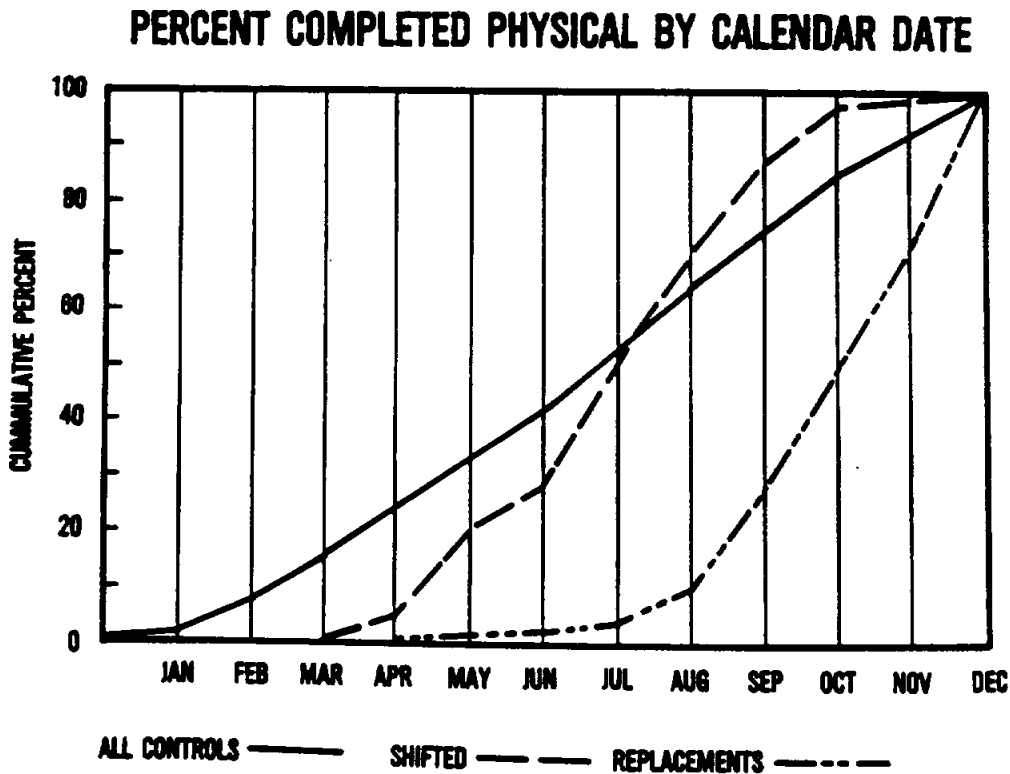


Figure V-4



Figures V-3 and V-4 show the cumulative percent of the Ranch Hand and comparison groups (original, shifted and replacement) completing the physical examination by time. Figure V-3 shows the similar time pattern of the Ranch Hand and original comparison group completing the physical examination. Figure V-4 shows that the shifted and replacement comparison groups were restricted in scheduling by the nature of the implementation of the design and contract time limitations. The overall comparison group cumulative completion of physical examination by calendar date is shown on both Figure V-3 and V-4. Fifty percent of the Ranch Hands and the original comparisons had completed their physical in May 1982, 50% of the shifted group had completed in July 1982, while 50% of the replaced group did not complete until October 1982.

7. Bias Assessment of Replacement Comparisons

From the above discussions and that in Chapters II and III, 2 questions are forthcoming which are of interest to inferential reliability. First, "Are the shifted and replaced comparisons valid for use without special statistical treatment?" Secondly, "What is the bias, if any, associated with the differential compliance to the physical examination?" The following sections deal with these 2 questions in turn.

8. Evaluation of the Replacement Comparison Participants.

Since the replacements used in the study, whether S, R or A, were simply the next individual in the randomized match set involved, the appropriate test for replacement bias is the test for O, S, R or A group differences while conditioning on the variables of age, occupation and race. Specifically, if S, R and A are unbiased groups they should appear to be random samples drawn from the same population as yielded the original (O) set, after adjustment for matching variables.

Tests of replacements against original comparisons were accomplished in accordance with procedures set out in the Study Protocol. Following the protocol, replacements for comparisons were tested first in terms of 3 primary variables to be ascertained on all participants: (a) subjective health assessment, (b) current utilization of long-term health care, and (c) recent work absenteeism pattern.

Statistical testing of these 3 primary variables and of additional questionnaire and physical examination variables was done in a prespecified manner. First, group A was tested against group R to determine if these groups could be combined. If R and A could be combined, the R + A group was tested against group S to determine if these groups could be combined. If R + A and S groups could be combined, O was tested against R + A + S. All testing was done at the 0.05 level. If the test for combination was not met at any stage, appropriate subtesting was performed. When the dependent variable was categorical, testing was performed with log-linear models adjusting by occupational category and age, with age dichotomized as less than 40 years and greater than or equal to 40 years providing groups of roughly equal sizes across occupational categories. When the dependent variable was continuous, analysis was performed with a general linear models program adjusting for occupational category and age as with the log-linear models. All of this testing was done to ascertain whether the S, R and A groups could be viewed as drawn from the same population as yielded the O group. Thus, the problem is one of hypothesis testing. Careful estimation of the magnitude or directionality of effects noted was not attempted. However, the reader can evaluate magnitude by reviewing data presented in the following paragraphs.

In reporting their health status, participants were allowed to use the categories: "excellent," "good," "fair" and "poor." Because of small sample sizes, the "fair" and "poor" responses were combined in the analysis of the data. Table V-10 provides a view of the data, collapsed across occupational

categories and age. No statistically significant differences between the S, R and A groups were found in either the partially compliant or fully compliant groups. However, when taken together, the fully compliant S, R and A groups appeared statistically different from the fully compliant original comparisons ($P < 0.001$). Additionally, the fully compliant O and S groups were found to be statistically different ($P = 0.01$), as were the fully compliant O and R groups ($P = 0.0045$). No statistically significant differences were noted among those individuals who took the questionnaire only.

Table V-10

SELF-ASSESSMENT OF HEALTH STATUS
(NON-BLACK PARTICIPANTS ONLY)

Status → Group ↓	Participants Who Took Questionnaire Only				Participants Who Took Questionnaire & Physical Examination			
	Excellent	Good	Fair or Poor	N	Excellent	Good	Fair or Poor	N
O	50.9%	34.7%	14.5%	173	38.0%	48.0%	14.0%	727
S	61.8%	26.5%	11.8%	34	36.4%	40.3%	23.4%	154
R	51.3%	38.2%	10.5%	76	49.6%	34.3%	16.1%	242
A	-	-	-	0	46.7%	43.3%	10.0%	30
Ranch Hand	52.5%	36.4%	11.0%	118	38.4%	41.4%	20.2%	976

O = Original Comparison
S = Shifted Comparison
R = Replacement Comparison
A = Air Force Interviewed Comparison

Use of long-term health care was assessed by inquiring about regular use of medications for heart, kidney, thyroid, renal and other disease states. No statistically significant differences were found between the O, S, R and A groups regarding regular use of medications. Table V-11 provides a view of the data collapsed across occupational categories and age.

Table V-11

MEDICATION USE
(NON-BLACK PARTICIPANTS ONLY)

Group	Participants Who Took Questionnaire Only (PC)		Participants Who Took Questionnaire and Physical Examination (FC)	
	Percent with Chronic Medication Use	N	Percent with Chronic Medication Use	N
O	23.6%	174	28.3%	728
S	14.7%	34	27.9%	154
R	19.7%	76	30.2%	242
A	-	0	16.7%	30
Ranch Hand	14.4%	118	29.4%	979

O = Original Comparison
 S = Shifted Comparison
 R = Replacement Comparison
 A = Air Force Interviewed Comparison

Work absenteeism was assessed by a consideration of reported time loss from work during the 6 months prior to interview. No statistically significant differences were noted between the O, S, R and A group on this parameter (relevant data provided in Table V-12).

Table V-12

WORK LOSS
(NON-BLACK PARTICIPANTS ONLY)

Group	Participants Who Took Questionnaire Only (PC)		Participants Who Took Questionnaire and Physical Examination (FC)	
	Percent with Work Loss	N	Percent with Work Loss	N
O	16.8%	173	20.5%	707
S	14.7%	34	21.1%	152
R	12.0%	75	18.6%	237
A	-	0	23.3%	30
Ranch Hand	18.8%	112	20.3%	955

O = Original Comparison
 S = Shifted Comparison
 R = Replacement Comparison
 A = Air Force Interviewed Comparison

Thus, for the 3 basic variables emphasized for test by the study protocol, the replacement comparisons (S+R+A) were found to be statistically significantly dissimilar from the originals on 1 variable, self-assessment of health. To more fully assess replacement-original differences, 9 additional variables from the questionnaire were examined: (1) household income, (2) participant education (high school or less, greater than high school), (3) participant anger scale, (4) participant psychoneurological erosion scale, (5) participant anxiety scale, (6) participant depression, (7) reported liver ailments, (8) spouse miscarriage rate, and (9) occurrence of acne. The fully compliant non-Black replacements (S+R+A) were observed to be statistically significantly different from the fully compliant original comparison participants as regards education ($P = 0.04$), anxiety level ($P = 0.02$), and psychoneurological erosion ($P = 0.02$). With respect to education 48.8% of the fully compliant replacement comparisons report more than a high school education, while 43.7% of the original comparisons report more than a high school education. Original fully compliant comparisons reported more moderate to severe anxiety than did the replacements (56.9% versus 55.6% respectively). Reported psychoneurological erosion addresses difficulties with mental tasks such as arithmetic work. The replacement comparisons reported erosion more commonly (37.2%) than did the original comparisons (30.2%). These measures of psychological status were not validated as truly measuring their intended end points and they are not necessarily statistically independent of one another, nonetheless, a picture of differences between the comparisons subsets is evident.

Thus, of 12 variables drawn from the questionnaire, 4 variables (reported health status, education, anxiety level and psychoneurological erosion) distinguish the replacement comparisons (S+R+A) from the original comparisons testing

at the 0.05 level. The differences observed are not only statistically significant but may also reflect clinically meaningful differences if the self-reporting is accurate. Analyses of bias have also been conducted using physical examination data end points to obtain a firmer evaluation, and these analyses are described in the following paragraphs.

Five laboratory variables have also been examined for evidence of differences among the comparison groups: white blood cell count (WBC), hemoglobin concentration (HGB), total bilirubin (TBIL), serum glutamic oxalic transaminase (SGOT) and lactic dehydrogenase (LDH). This testing is summarized in Table V-13. The analyses were performed with a general linear models program, operating on WBC and HGB in natural units and TBIL, SGOT and LDH in logarithmic units. It is clear from Table V-13 that there is definite indication of comparison group differences.

Table V-13

SUMMARY OF BIAS ASSESSMENTS OF REPLACEMENT
COMPARISONS USING LABORATORY MEASURES
(NON-BLACK PARTICIPANTS ONLY)

<u>Clinical Variable</u>	<u>Adjusted Mean For Original (O) Comparisons</u>	<u>Adjusted Mean For All Replacements (S+R+A)</u>	<u>P Value For Mean Differential</u>
WBC	7.24	7.78	0.027
HGB	16.0	15.9	0.522
TBIL	0.577	0.609	0.063
SGOT	33.1	32.7	0.498
LDH	142.0	141.2	0.265

Lastly, 13 clinical variables from the physical examination itself were evaluated for O, S, R, A comparison group differences. As summarized in Table V-14, statistically significant differences were found.

Table V-14

SUMMARY OF BIAS ASSESSMENTS OF REPLACEMENT COMPARISONS
USING MEASURES FROM THE PHYSICAL EXAMINATION

*1. Systolic Blood Pressure	No differences detected
*2. Diastolic Blood Pressure	No differences detected
*3. Posterior Tibial Pulse	{S statistically different from R + A O statistically different from R + A
*4. Dorsalis Pedis Pulse	No differences detected
*5. EKG	{S statistically different from R + A O statistically different from R + A
6. Vibration Sense	{S statistically different from R + A O not different from R + S + A
7. Tremor	{S statistically different from R + A O statistically different from S
8. Nerve conduction velocity above the elbow	No differences detected
9. Nerve conduction velocity below the elbow	No differences detected
10. Peroneal nerve conduction velocity	No differences detected
11. Full Scale Intelligence Quotient	No differences detected
12. MMPI Scale D	{S statistically different from R + A O statistically different from R + A
13. MMPI Scale L	No differences detected

*Black participants removed.

Taken together the analyses described above imply very strongly that the S, R and A comparison groups are not random samples drawn from the same population as the original comparisons (O). Since the comparison group differences are not observed in all variables studied, a possible approach is to perform a prior test of significance (PTS) to test for appropriateness of replacement use, followed when possible by a Ranch Hand-all comparison test. This use of a PTS has been discussed with appreciable detail in the statistical literature (Bozivich et al, 1956; Bancroft, 1964; Kale and Bancroft, 1967; Arnold, 1970; Cohen, 1974). Recommendations in this literature suggest a preliminary test for combination using an alpha level of 0.25, followed by a test of differences at an alpha level of 0.05. Calculations of study power with and without

the PTS have indicated that, given the sample sizes in this study, the PTS only provides partial protection against inferential bias. This result can be understood by reference to Figure V-5 where 2 power curves are given.

Figure V-5

POWER CURVES FOR ALTERNATIVE ANALYTICAL METHODS

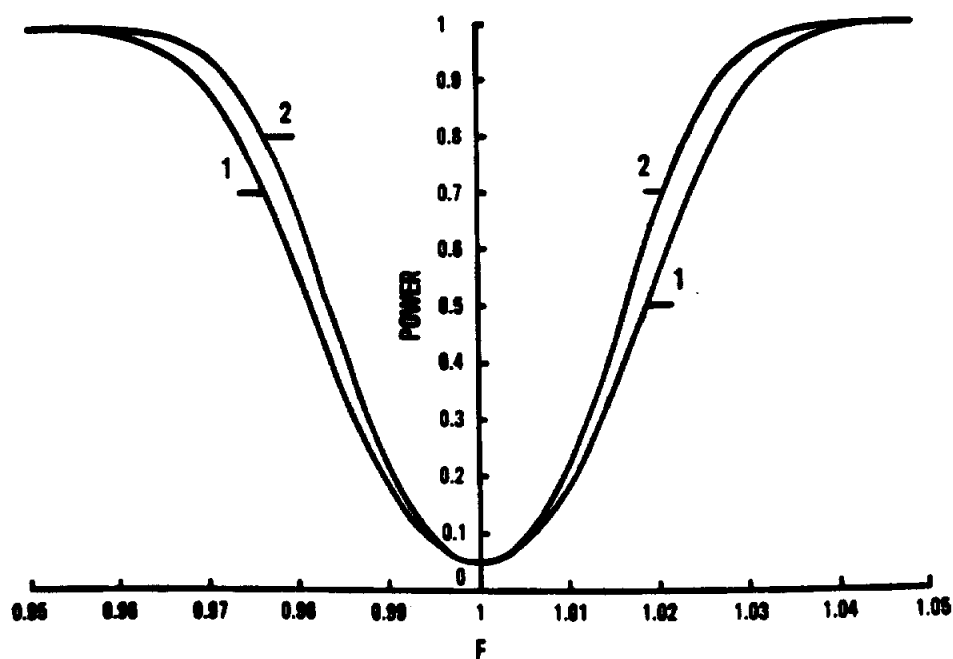


Figure V-5. Curve 1: Power curve for Ranch Hand-original comparison tests on means. Curve 2: Power curve for Ranch Hand-comparison tests on means assuming replacement comparisons are unbiased. F is the symbol for ratios of Ranch Hand-comparison means.

The lower power curve (curve #1) is for a test of difference between the Ranch Hand group ($N=1045$) and the original comparisons ($N=773$). The upper curve (curve #2) is for the same test of difference but between the Ranch Hand group and all comparisons ($N=1224$) assuming that the replacements are unbiased. These curves are drawn for a hypothetical clinical variable with ratio of standard deviation to mean being 0.200. The variable F is the ratio of the exposed mean to the comparison mean. The slight displacement of the 2 curves in the vertical direction (power) is easily negated by small degrees of bias in the replacement comparisons.

The Study Protocol reflects a strong concern for a variety of biases that may be operating in this study. The effect of the potential bias, by using the shifted and replacement members of the comparison group, was not uniformly viewed by the Principal Investigators. Because of time constraints, the Science Panel was not convened to address this complex issue. Instead, a management decision was made to base the primary clinical analyses upon a contrast of the Ranch Hand group and members of the original comparison group. For completeness of data descriptions, some chapters additionally contain analyses founded upon the entire comparison group.

9. Noncompliance Bias

The data in the previous section suggest that a degree of self-selection did occur in association with compliance to the physical examination, indicating that the group who came to physical examination may be biased from the original sample. Since this report emphasizes analysis of data from fully compliant participants, selection biases associated with physical examination compliance are of importance. Table V-15 displays differences between fully and partially compliant study participants.

Table V-15

DIFFERENCES BETWEEN FULLY COMPLIANT
(TOOK QUESTIONNAIRE AND PHYSICAL EXAMINATION)
AND PARTIALLY COMPLIANT (TOOK QUESTIONNAIRE) STUDY PARTICIPANTS:
P VALUES FOR TEST OF NO DIFFERENCE

	<u>Ranch Hand Fully Compliant Versus Partially Compliant</u>	<u>Original Comparison Fully Compliant Versus Partially Compliant</u>
Health Status	0.006	0.004
Medication Use	<0.001	0.23
Work Loss	0.79	0.30
Household Income	0.32	0.86
Education	0.66	0.39
Anger	<0.001	0.01
Anxiety	0.020	0.61
Erosion	<0.001	0.002
Depression	0.007	0.36
Liver Ailments	0.76	0.64
Miscarriages	0.97	0.077
Acne	0.37	0.75

Eighty-seven percent of the Ranch Hand group were compliant to the physical examination while 76% of the original eligible comparisons attended. Let RR_{obs} be the observed relative risk calculated from the physical examination data and RR be the actual relative risk of the originally drawn groups. Direct algebraic considerations provide the relationship

$$RR = \frac{0.13 \gamma_e + 0.87}{0.24 \gamma_c + 0.76} \quad RR_{obs} \quad \text{Equation \#1}$$

In this equation, γ_e is the ratio of the prevalence of the finding in the Ranch Hand group noncompliant to physical examination, to the prevalence in Ranch Hand individuals who were examined; the term γ_c is the same ratio for the comparison group. In other words, the values γ_e and γ_c are within-group noncompliant-to-compliant relative risks. The values of γ_e and γ_c are in fact not known so that RR can in fact not be known with exactness. Were $RR_{obs} = 1.00$ and were the finding rate 0.100 in the fully compliant comparison group, γ_e and γ_c could both range from zero to 10, indicating that RR could take values from 0.28 to 2.86. Thus, noncompliance to the physical examination is a serious concern in the attempt to properly infer herbicide effects from group differences noted at physical examination.

It is possible to develop an indication of the magnitude of the within-group relative risks γ_e and γ_c using data from the questionnaire. From Table V-15, it is clear that in several instances (roughly 50%) the fully compliant replacements are not statistically different from the partially compliant or, approximately, $\gamma_e = \gamma_c = 1.0$. In these cases, an observed relative risk, RR_{obs} , is at least approximately equal to the actual relative risk, RR of the original sample. On the other hand, using the health status data, γ_e is estimated to be 0.54 while γ_c is 1.04 for the categories "fair-poor" health, indicating (using Equation #1) that $RR = 0.93 RR_{obs}$. This result implies the possibility that the use of physical examination data can overestimate a relative risk by 7%. On the other hand, for the erosion scale γ_e is 0.52, while γ_c is 0.63, providing $RR = 1.03 RR_{obs}$, which implies the possibility that the physical examination could underestimate relative risk by 3%.

These calculations of γ_e and γ_c use questionnaire data, and thus, the results are indications only of bias in the physical examination, due to the extrapolation from 1 data set to another. Nevertheless, the results do indicate a range of bias which is much smaller than the range obtained when no assumptions about γ_e and γ_c are made.

It is difficult to conceive of a partially compliant rate or proportion as being different from a fully compliant rate or proportion by more than a factor of 2. Thus it may be assumed that

$$0.5 \leq \gamma_e \leq 2.0$$

$$0.5 \leq \gamma_c \leq 2.0$$

under this assumption

$$0.75 RR_{obs} \leq RR \leq 1.28 RR_{obs}$$

An inequality such as the one above should be applied to each study result reported here to reflect the possible effect of selection bias. If the above inequality is used, the smallest observed relative risk that can be considered actually larger than 1 is 1.33 ($=0.75^{-1}$) and the largest observed relative risk that can be considered actually smaller than 1 is 0.78 ($=1.28^{-1}$). Or, as a simpler rule of thumb, full sample relative risks may be assumed to be within $\pm 30\%$ of observed relative risks. Of course, this measure of uncertainty due to noncompliance must be added to the uncertainty due to finite sample sizes, and to other sources of possible inferential error.

It is not feasible to numerically evaluate the degree of bias in physical examination measurements of continuously distributed variables such as blood pressure, hemoglobin concentration or pulmonary volumes, using questionnaire data, as no analogous values were obtained from the questionnaire. An equation similar to Equation #1 holds for the ratio of group mean values for a continuous variable, namely:

$$RAT = \frac{0.13 \gamma_e^1 + 0.87}{0.24 \gamma_c^1 + 0.76} RAT_{obs} \quad \text{Equation \#2}$$

In this equation, RAT_{obs} is the ratio of the Ranch Hand fully compliant mean to the comparison fully compliant mean, RAT is the ratio of the means of the complete original samples, γ_e^1 is the ratio of the partially compliant mean to the fully compliant mean in the Ranch Hand set and γ_c^1 is the same ratio for the comparison participants. Estimates of γ_e^1 and γ_c^1 are not available; however, it is difficult to conceive of a partially compliant mean as different from a fully compliant mean in the same group by more than 20%; whence, we assume:

$$\begin{aligned} 0.80 &\leq \gamma_e^1 \leq 1.20 \\ 0.80 &\leq \gamma_c^1 \leq 1.20 \end{aligned}$$

Under this assumption

$$0.93 RAT_{obs} \leq RAT \leq 1.08 RAT_{obs}$$

that is, full sample ratios are anticipated to be within $\pm 8\%$ of observed sample ratios of means. The potential error in sample mean ratios portrayed above must be considered by the reader in the interpretation of mean shift data presented in this report.

10. Summary and Conclusion

The comparison group in this study is divisible into 3 subgroups: original comparisons, shifted comparisons and replacements. Due to study implementation and contractual constraints, the shifted and replaced comparison groups were scheduled differently from the original comparison group for the study questionnaire and physical examination. The original comparisons were handled in a manner essentially identical to that of the Ranch Handers.

Analysis has shown that replacements differ from original comparisons on compliance to questionnaire and physical examination; however, shifted comparisons are not statistically significantly different from originals on these parameters. Both shifted and replacement comparisons have been found to be statistically significantly different from the original comparisons on a variety of questionnaire and physical examination measures. This source of potential bias is completely avoided in this report through the primary use of the original comparisons in hypothesis testing.

Differential compliance to the physical examination occurred with 87% of the Ranch Handers and 76% of the comparisons attending. This fact raises the concern for a second bias which cannot be avoided, and it could be a result of media and Ranch Hand Association support for this study. It is suggested, however, that this bias is not large. Worst-case estimates imply that observed relative risks are displaced from correct relative risks by no more than 30% by noncompliance effects, and observed mean ratios are displaced by no more than 8%.